

# JOINT DICTIONARY LEARNING FOR EXAMPLE-BASED IMAGE SUPER-RESOLUTION

Mojtaba Sahraee-Ardakan<sup>1</sup>, Mohsen Joneidi<sup>2</sup>

<sup>1</sup>Department of Electrical Engineering, Sharif University of Technology, Iran.

<sup>2</sup>Department of Electrical Engineering, Ferdowsi University of Mashhad, Iran

## ABSTRACT

In this paper, we propose a new joint dictionary learning method for example-based image super-resolution (SR), using sparse representation. The low-resolution (LR) dictionary is trained from a set of LR sample image patches. Using the sparse representation coefficients of these LR patches over the LR dictionary, the high-resolution (HR) dictionary is trained by minimizing the reconstruction error of HR sample patches. The error criterion used here is the mean square error. In this way we guarantee that the HR patches have the same sparse representation over HR dictionary as the LR patches over the LR dictionary, and at the same time, these sparse representations can well reconstruct the HR patches. Simulation results show the effectiveness of our method compared to the state-of-art SR algorithms.

**Index Terms**— Image super-resolution, sparse representation, dictionary learning

## 1. INTRODUCTION

Super-resolution is the problem of reconstructing a high resolution<sup>1</sup> image from one or several low resolution images [1]. It has many potential applications like enhancing the image quality of low-cost imaging sensors (e.g., cell phone cameras) and increasing the resolution of standard definition (SD) movies to display them on high definition (HD) TVs, to name a few.

Prior to SR methods, the usual way to increase resolution of images was to use simple interpolation-based methods such as bilinear, bicubic and more recently the resampling method described in [2] among many others. However all these methods suffer from blurring high-frequency details of the image especially for large upscaling factors (the amount by which the resolution of image is increased in each dimension). Thus, over the last few years, a large number of SR algorithms have been proposed [3]. These methods can be classified into two categories: multi-image SR, and single-image SR.

Since the seminal work by Tsai and Huang [4] in 1984, many multi-image SR techniques were proposed [5, 6, 7, 8]. In the conventional SR problem, multiple images of the same scene with subpixel motion are required to generate the HR

image. However the performance of these SR methods are only acceptable for small upscaling factors (usually smaller than 2). As the upscaling factor increases, the SR problem becomes severely ill-conditioned and a large number of LR images are needed to recover the HR image with acceptable quality.

To address this problem, example-based SR techniques were developed which require only a single LR image as input [9]. In these methods, an external training database is used to learn the correspondence between manifolds of LR and HR image patches. In some approaches, instead of using an external database, the patches extracted from the LR image itself across different resolutions are used [10]. In [9] Freeman *et al.* used a Markov network model for super-resolution. Inspired by the ideas in locally linear embedding (LLE) [11], the authors of [12] used the similarity between manifolds of HR patches and LR patches to estimate HR image patches. Motivated by results of compressive sensing [13], Yang *et al.* in [14] and [15] used sparse representation for SR. In [16] they introduced coupled dictionary training in which the sparse representation of LR image patches better reconstructs the HR patches.

Recently, joint and coupled learning methods are utilized for efficient modeling of correlated sparsity structures [17, 15]. However joint learning methods and the coupled learning methods proposed in [14, 15, 18] still does not guarantee that the sparse representation of HR image patches over the HR dictionary is the same as the sparse representation of LR patches over LR dictionary. To address this problem, in this paper we propose a direct way to train the dictionaries that enforces the same sparse representation for LR and HR patches. Moreover since the HR dictionary is trained by minimizing the final error in reconstruction of HR patches, the reconstruction error in our method is smaller.

The rest of this paper is organized as follows. In section 2, Yang's method for super-resolution via sparse representation is reviewed. In section 3, a flaw in Yang's method is discussed, and our method to solve this problem is presented. Finally, section 4 is devoted to simulation results.

<sup>1</sup>In this article by resolution we mean spatial resolution.

## 2. REVIEW OF SUPER-RESOLUTION VIA SPARSE REPRESENTATION

In SR via sparse representation we are given two sets of training data: a set of LR image patches, and a set of corresponding HR image patches. In other words, in the training data we have pairs of LR and HR image patches. The goal of SR is to use this database to increase the resolution of a given LR image.

Let  $\{\mathbf{y}_i\}_{i=1}^N$  be the set of LR patches (each patch is arranged into a column vector  $\mathbf{y}_i$ ) and  $\{\mathbf{x}_i\}_{i=1}^N$  be the set of corresponding HR patches. In SR using sparse representation, the problem is to train two dictionaries  $\mathbf{D}_l$  and  $\mathbf{D}_h$  for the set of LR patches (or a feature of these patches) and HR patches respectively, such that for any LR patch  $\mathbf{y}_i$ , its sparse representation  $\mathbf{w}_i$  over  $\mathbf{D}_l$ , reconstructs the corresponding HR patch  $\mathbf{x}_i$  using  $\mathbf{D}_h$ :  $\mathbf{x}_i \approx \mathbf{D}_h \mathbf{w}_i$  [15]. Towards this end, first the dictionary learning problem is briefly reviewed in section 2.1. Then the dictionary learning method for SR proposed in [15] is studied in section 2.2. Finally in section 2.3, it is shown how these trained dictionaries can be used to perform SR on a LR image.

### 2.1. Dictionary learning

Given a set of signals  $\{\mathbf{x}_i\}_{i=1}^N$ , dictionary learning is the problem of finding a wide matrix  $\mathbf{D}$  over which the signals have sparse representation [19]. This problem is highly related to subspace identification [20]. However, sparsity helps us to turn the subspace recovery to a well-defined problem. This approach has attracted lot of attentions in the last decade and found diverse applications [21, 22, 23]. If we denote the sparse representation of  $\mathbf{x}_i$  over  $\mathbf{D}$  by  $\mathbf{w}_i$ , the dictionary learning problem can be formulated as

$$\min_{\mathbf{D}, \{\mathbf{w}_i\}_{i=1}^N} \sum_{i=1}^N \|\mathbf{w}_i\|_0 \quad s.t. \|\mathbf{x}_i - \mathbf{D}\mathbf{w}_i\|_2^2 \leq \epsilon, i = 1, \dots, N \quad (1)$$

in which the  $\|\cdot\|_0$  is the  $l_0$ -norm which is the number of nonzero components of a vector and  $\epsilon$  is a small constant which determines the maximum tolerable error in sparse representations. Replacing the  $l_0$ -norm by  $l_1$ -norm, Yang *et al.* in [15] used the following formulation for sparse coding instead of (1)

$$\min_{\mathbf{D}, \{\mathbf{w}_i\}_{i=1}^N} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{D}\mathbf{w}_i\|_2^2 + \lambda \sum_{i=1}^N \|\mathbf{w}_i\|_1. \quad (2)$$

By defining  $\mathbf{X} \triangleq [\mathbf{x}_1 \cdots \mathbf{x}_N]$  and  $\mathbf{W} \triangleq [\mathbf{w}_1 \cdots \mathbf{w}_N]$ , it can be rewritten in matrix form as

$$\min_{\mathbf{D}, \mathbf{W}} \|\mathbf{X} - \mathbf{D}\mathbf{W}\|_F^2 + \lambda \|\mathbf{W}\|_1, \quad (3)$$

in which  $\|\cdot\|_F$  stands for the Frobenius norm. (2) and (1) are not equivalent, but closely related. (2) can be interpreted

as minimizing the representation error of signals over the dictionary, while forcing these representations to be sparse by adding a  $l_1$ -regularization to the error. Therefore  $\lambda$  can be used as a parameter that balances the sparsity and the error; a larger  $\lambda$  results in sparser representations with larger errors.

### 2.2. Dictionary learning for SR

Given the sets of LR and HR training patches,  $\{\mathbf{y}_i\}_{i=1}^N$  and  $\{\mathbf{x}_i\}_{i=1}^N$ , by defining  $\mathbf{Y} \triangleq [\mathbf{y}_1 \cdots \mathbf{y}_N]$ , and having (3) in mind, Yang *et al.* in [15] proposed the following joint dictionary learning to ensure that the sparse representation of LR patches over  $\mathbf{D}_l$  is the same as sparse representation of HR image patches over  $\mathbf{D}_h$ :

$$\min_{\mathbf{D}_l, \mathbf{D}_h, \mathbf{W}} \|\mathbf{Y} - \mathbf{D}_l \mathbf{W}\|_F^2 + \|\mathbf{X} - \mathbf{D}_h \mathbf{W}\|_F^2 + \lambda \|\mathbf{W}\|_1 \quad (4)$$

The key point here is that they have used the same matrix  $\mathbf{W}$  for sparse representation of both LR and HR patches to make sure that their representation is the same over the dictionaries  $\mathbf{D}_l$  and  $\mathbf{D}_h$ . If we define the concatenated space of HR and LR patches:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{X} \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \mathbf{D}_l \\ \mathbf{D}_h \end{bmatrix}$$

then joint dictionary training (4) can also be written equivalently as

$$\min_{\mathbf{D}, \mathbf{W}} \|\mathbf{Z} - \mathbf{D}\mathbf{W}\|_F^2 + \lambda \|\mathbf{W}\|_1. \quad (5)$$

This formulation is clearly the same as (3). In other words, in the concatenated space, joint dictionary learning is the same as conventional dictionary learning, and any dictionary learning algorithm can be used for joint dictionary learning.

### 2.3. Super-Resolution

After training the two dictionaries  $\mathbf{D}_l$  and  $\mathbf{D}_h$ , the input LR image can be super-resolved using the following steps:

1. The input LR image is divided into a set of overlapping LR patches:  $\{\mathbf{y}_i^{LR}\}_{i=1}^M$ .

2. From each image patch  $\mathbf{y}_i^{LR}$ , subtract its mean,  $\eta_i$ ,

$$\hat{\mathbf{y}}_i^{LR} \triangleq \mathbf{y}_i^{LR} - \eta_i,$$

and find its sparse representation over  $\mathbf{D}_l$

$$\mathbf{w}_i = \arg \min_{\boldsymbol{\alpha}_i} \|\hat{\mathbf{y}}_i^{LR} - \mathbf{D}_l \boldsymbol{\alpha}_i\|_2^2 + \lambda \|\boldsymbol{\alpha}_i\|_1.$$

3. Using the sparse representation of each LR patch and its mean, the corresponding HR patch is estimated by

$$\hat{\mathbf{x}}_i^{HR} = \mathbf{D}_h \cdot \mathbf{w}_i, \quad \mathbf{x}_i^{HR} = \hat{\mathbf{x}}_i^{HR} + \eta_i.$$

4. Combining the estimated HR image patches, the output HR image is generated.

### 3. OUR PROPOSED METHOD

Our method for SR is to improve the dictionary learning part of Yang's method, described in section 2.2. Having the dictionaries trained, the rest of the method is the same as what described in section 2.3.

As mentioned earlier, in SR the dictionaries should be trained in a way that the sparse representation of each LR patch well reconstructs the corresponding HR patch. The Yang's method uses (4) to accomplish this. It uses the same sparse representation matrix  $\mathbf{W}$  for both LR and HR patches to ensure that each LR and HR patch, both have the same sparse representation. However as can be seen from (5), this joint dictionary learning is only optimal in the concatenated space of LR and HR patches, but if we look at the space of LR and HR patches separately, we may find a sparser representation for some patches than the sparse representation found in the concatenated space.

To address this problem, note first that the SR method described in section 2.3 consists of two distinct operations: finding the sparse representation of the LR patch, and the reconstruction of the HR patch. Then, we note that the first operation uses only  $\mathbf{D}_l$ , and the second operation uses only  $\mathbf{D}_h$ . Therefore, instead of training the dictionaries jointly as in (4), we propose to train  $\mathbf{D}_l$  for LR patches solely, and then to train the HR dictionary by minimizing the reconstruction error when sparse representation of LR patches are used.

Mathematically, we propose to train the LR dictionary as

$$\min_{\mathbf{D}_l, \mathbf{W}} \|\mathbf{Y} - \mathbf{D}_l \mathbf{W}\|_F^2 + \lambda \|\mathbf{W}\|_1, \quad (6)$$

which is a conventional dictionary learning problem. After training the LR dictionary, for each LR patch, its sparse representation  $\mathbf{w}_i$  is found over  $\mathbf{D}_l$  (note that this step is already done during the dictionary training in (6))

$$\mathbf{W} = \underset{\mathbf{V}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{D}_l \mathbf{V}\|_F^2 + \lambda \|\mathbf{V}\|_1. \quad (7)$$

Using the sparse representation of LR patches  $\mathbf{W}$ , the HR dictionary  $\mathbf{D}_h$  is found such that the reconstruction error of the corresponding HR patches are minimized, that is,

$$\mathbf{D}_h = \underset{\mathbf{D}_h}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{D}_h \mathbf{W}\|_F^2. \quad (8)$$

This is an unconstrained quadratic optimization problem which has the following closed-form solution:

$$\mathbf{D}_h = \mathbf{X} \mathbf{W} (\mathbf{W} \mathbf{W}^T)^{-1} = \mathbf{X} \mathbf{W}^\dagger \quad (9)$$

in which  $(\cdot)^T$  and  $(\cdot)^\dagger$  represent transpose and pseudo-inverse of a matrix, respectively.

Note that unlike Yang's method, in the proposed method  $\mathbf{D}_h$  is not trained in a way that explicitly enforces the sparsity of representation of HR patches over it, rather it is trained to minimize the final reconstruction error.



**Fig. 1.** Results of Lena image magnified by a factor of 2 using: (b) Bicubic interpolation, (c) Yang's method, (d) our proposed method. The original image is also given in (a) for comparison.

### 4. SIMULATION RESULTS

In this section we compare the performance of our method with Yang's method. The error criteria used here are Peak Signal to Noise Ratio (PSNR) and Structural SIMilarity (SSIM) index [24]. PSNR criterion is defined as

$$\text{PSNR} = 20 \log_{10} \left( \frac{255}{\sqrt{\text{MSE}}} \right), \quad (10)$$

where MSE is mean square error given by

$$\text{MSE} = \frac{\|\mathbf{I}_{SR} - \mathbf{I}_g\|_F^2}{mn}, \quad (11)$$

in which  $\mathbf{I}_g$  is the original distortion-free image, and  $\mathbf{I}_{SR}$  is the super-resolved image derived from the SR algorithm, and  $m$  and  $n$  are dimensions of the image in pixels.

For the definition of SSIM refer to [24]. From these definitions it is clear that higher PSNR means less mean square error, however it does not necessarily mean a better image quality when perceived by human eye. Many other error criteria have been proposed to solve this problem of PSNR. SSIM is one of these error criteria. But still PSNR is widely used because of its simple mathematical form. This can be seen in (8) where MSE is to train the HR dictionary, but in order to show the effectiveness of our method, here we use both SSIM

**Table 1.** PSNR and SSIM of some images magnified using bicubic interpolation, Yang’s method and our proposed method. The average PSNRs and SSIMs are given in the last row. Best performance in each row is written in boldface.

		Bicubic	Yang	proposed
Lena	PSNR	32.79	34.73	<b>34.86</b>
	SSIM	0.9012	0.9268	<b>0.9283</b>
Parthenon	PSNR	26.50	27.77	<b>27.89</b>
	SSIM	0.8334	0.8737	<b>0.8762</b>
Baboon	PSNR	24.66	25.30	<b>25.39</b>
	SSIM	0.9529	0.9872	<b>0.9873</b>
Barbara	PSNR	27.93	<b>28.61</b>	28.59
	SSIM	0.9609	<b>0.9852</b>	<b>0.9852</b>
Flower	PSNR	30.51	33.24	<b>33.36</b>
	SSIM	0.9230	0.9526	<b>0.9538</b>
<b>Average</b>	PSNR	28.47	29.93	<b>30.02</b>
	SSIM	0.9143	0.9451	<b>0.9462</b>

and PSNR to compare images produced by our method with Yang’s method.

To make a fair comparison, the same set of 80000 training data patches sampled randomly from natural images is used to train dictionaries for both Yang’s and our method. The size of LR patches is  $5 \times 5$  and they are magnified by a factor of 2, i.e. the size of generated HR image patches is  $10 \times 10$ . The LR patches extracted from the input image have a 4 pixel overlap. Dictionary size is fixed at 1024, and  $\lambda = 0.15$  is used for both methods as in [15].

In Fig. 1, simulation results of Yang’s method and proposed method on Lena image can be seen. The original image and the image magnified using bicubic interpolation are also given as references. The PSNRs of these images are 32.79dB, 34.73dB and 34.86dB for bicubic interpolation, Yang’s method and ours respectively. It is clear that the quality of images magnified by SR is much better than the image magnified by bicubic interpolation and the details are more visible, which has resulted in sharper images. But the difference between image (c) and image (d) is not noticeable visually, although the PSNR of image (d) which is super-resolved by our method is about 0.1dB higher.

In Table 1 the PSNRs and SSIMs of some images produced by our method is compared with those of Yang’s method and bicubic interpolation. Almost all of the images recovered by our method have higher PSNRs than images recovered by Yang’s method. The average PSNRs given in the last row show that our method performs slightly better than Yang’s method on average.

The SSIMs in Table 1 also confirm that our method is

performing better than Yang’s method. The images super-resolved by the proposed method have on average a higher SSIM than images recovered by Yang’s method. Since SSIM is much more consistent with the image quality as it is perceived by human eye compared to PSNR, higher SSIM of images recovered by our method suggests that they also have better visual quality.

## 5. CONCLUSION AND FUTURE WORKS

In this paper, we presented a new dictionary learning algorithm for example-based SR. The dictionaries were trained from a set of sample LR and HR image patches in order to minimize the final reconstruction error. Simulation results on real images showed the effectiveness of our algorithm in super-resolving images with less error compared to Yang’s method. The average PSNR and average SSIM of images produced by our method were higher than images recovered by Yang’s method. In future, we can extend this work by training the HR dictionary using a better error criterion instead of PSNR. One of the advantages of our method is that training of  $D_h$  is separated from  $D_l$  in (6) and (8). We can use another error criterion that better represents the image quality like SSIM in (8) without making the training of  $D_l$  more complex. Changing the error criterion in each of Yang’s methods will make the optimizations in their algorithms much more complex.

## 6. REFERENCES

- [1] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang, “Super-resolution image reconstruction: a technical overview,” *Signal Processing Magazine, IEEE*, vol. 20, no. 3, pp. 21–36, 2003.
- [2] Lei Zhang and Xiaolin Wu, “An edge-guided image interpolation algorithm via directional filtering and data fusion,” *Image Processing, IEEE Transactions on*, vol. 15, no. 8, pp. 2226–2238, 2006.
- [3] Peyman Milanfar, *Super-resolution imaging*, CRC Press, 2010.
- [4] RY Tsai and T.S. Huang, “Multiframe image restoration and registration,” *Advances in computer vision and Image Processing*, vol. 1, no. 2, pp. 317–339, 1984.
- [5] Michael Elad and Arie Feuer, “Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images,” *Image Processing, IEEE Transactions on*, vol. 6, no. 12, pp. 1646–1658, 1997.
- [6] Michael Elad and Arie Feuer, “Super-resolution reconstruction of image sequences,” *Pattern Analysis and*

- Machine Intelligence, IEEE Transactions on*, vol. 21, no. 9, pp. 817–834, 1999.
- [7] David Capel and Andrew Zisserman, “Super-resolution from multiple views using learnt image models,” in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. IEEE, 2001, vol. 2, pp. II–627.
  - [8] Sina Farsiu, M Dirk Robinson, Michael Elad, and Peyman Milanfar, “Fast and robust multiframe super resolution,” *Image processing, IEEE Transactions on*, vol. 13, no. 10, pp. 1327–1344, 2004.
  - [9] William T Freeman, Thouis R Jones, and Egon C Pasztor, “Example-based super-resolution,” *Computer Graphics and Applications, IEEE*, vol. 22, no. 2, pp. 56–65, 2002.
  - [10] Daniel Glasner, Shai Bagon, and Michal Irani, “Super-resolution from a single image,” in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 349–356.
  - [11] Sam T Roweis and Lawrence K Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
  - [12] Hong Chang, Dit-Yan Yeung, and Yimin Xiong, “Super-resolution through neighbor embedding,” in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. IEEE, 2004, vol. 1, pp. I–275.
  - [13] Emmanuel J Candès, “Compressive sampling,” in *Proceedings of the International Congress of Mathematicians: Madrid, August 22-30, 2006: invited lectures*, 2006, pp. 1433–1452.
  - [14] Jianchao Yang, John Wright, Thomas Huang, and Yi Ma, “Image super-resolution as sparse representation of raw image patches,” in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.
  - [15] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma, “Image super-resolution via sparse representation,” *Image Processing, IEEE Transactions on*, vol. 19, no. 11, pp. 2861–2873, 2010.
  - [16] Jianchao Yang, Zhaowen Wang, Zhe Lin, Scott Cohen, and Thomas Huang, “Coupled dictionary training for image super-resolution,” *Image Processing, IEEE Transactions on*, vol. 21, no. 8, pp. 3467–3478, 2012.
  - [17] A. Taalimi, H. Qi, and R. Khorsandi, “Online multi-modal task-driven dictionary learning and robust joint sparse representation for visual tracking,” in *Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on*, Aug 2015, pp. 1–6.
  - [18] A. Taalimi, A. Rahimpour, C. Capdevila, Z. Zhang, and H. Qi, “Robust coupling in space of sparse codes for multi-view recognition,” in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 3897–3901.
  - [19] Michael Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*, Springer, 2010.
  - [20] Mostafa Rahmani and George Atia, “Innovation pursuit: A new approach to subspace clustering,” *arXiv preprint arXiv:1512.00907*, 2015.
  - [21] Shervin Minaee, Amirali Abdolrashidi, and Yao Wang, “Screen content image segmentation using sparse-smooth decomposition,” in *2015 49th asilomar conference on signals, systems and computers*. IEEE, 2015, pp. 1202–1206.
  - [22] Mahdi Abavisani, Mohsen Joneidi, Shideh Rezaeifar, and Shahriar Baradaran Shokouhi, “A robust sparse representation based face recognition system for smartphones,” in *2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 2015, pp. 1–6.
  - [23] M Joneidi, M Rahmani, HB Golestani, and M Ghanbari, “Eigen-gap of structure transition matrix: A new criterion for image quality assessment,” in *Signal Processing and Signal Processing Education Workshop (SP/SPE), 2015 IEEE*. IEEE, 2015, pp. 370–375.
  - [24] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.